EXAM I REVIEW

Stat 120 | Fall 2025

Prof Amanda Luby

The notes that we go through in class are good "study guides" for the exam (including exercises, examples, and definitions). You don't need to memorize exact definitions or formulas, but you should have a solid idea of what terms mean.

Topics for Exam 1:

1. Collecting Data
    a) Cases and variables
    b) Population vs Sample
    c) Random Sampling
    d) Random Assignment
    e) Sampling Bias
2. EDA for categorical variables
    a) Proportions
    b) Frequency and two-way tables
    c) Joint, marginal, and conditional distributions
    d) Barplots
    e) Stacked barplots
3. EDA for quantitative variables
    a) Measures of center
    b) Histograms and dotplots
    c) Standard deviation
    d) Z-scores
    e) Five number summary and boxplots
    f) Outliers
4. Describing Relationships
    a) Correlation
    b) Linear Regression
    c) Influential points
    d) Residual plots
    e) Transformations for linear regression

Calculator: While you might be asked to do basic arithmetic (similar to HW problems), if you get the answer to a point you would type it into the calculator, that will be full credit. I will bring basic calculators that you can use if you wish, but you may not use the calculator on your phone or any calculator that can make graphs or be programmed.

R Questions: You won't be able to use your computer and won't be asked to write down any R code. I will not give you R code and ask you to tell me what it does. You will be asked to look at output (e.g. a graph or `lm()` output) and interpret the results.

Academic Integrity: I will ask you all to sign a statement that you have not received (and will not provide) any guidance or help on the exam problems, and will not discuss the exam with anyone inside or outside of class until it is returned to you. This includes using materials from prior versions of this class to study. Any academic integrity violations on the exam will be reported to the Provost's office.

Additional Practice: In addition to your notes, R activities in class, and homework; the "Unit A Essential Synthesis" on page 177 of the textbook provides a nice overview and some conceptual practice problems. *Note:* don't over-study these practice problems, since the format of the exam will be entirely different. Focus your time and energy on understanding notes and homework problems first.

Format of the exam: The exam will be a pseudo-data-analysis. The best way to describe it is to imagine an R activity from class where someone else did all of the coding and found all of the R output, and you are tasked with answering all of the substantive questions about the data.

The attached questions provide examples of the *format* of questions I will ask, using the dataset from Activity 6. They do not represent the full range of *content* or the *length* of the exam.